

SEQC2 data access from NCBI

Chunlin Xiao

Main ftp site

- <ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/>
- Initial 10T allocated
- Mainly for reads data for now based on the discussions between Wenming and NCBI leaders
- Just operational – starting to move data to the ftp site (HapMap trio fastqs already on ftp; Quartet fastqs pending on Leming's approval)

Data access policy for SEQC data

- Initial drafted by Weida, modified and agreed by others (Wenming, Huixiao, Chunlin etc...)
- Policy on main ftp site

ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/Confidentiality_Terms_for_Accessing_the_SEQC2_Datasets.txt

Confidentiality Terms for Accessing the SEQC2 Datasets

1. SEQC2 is the FDA-led Sequencing Quality Control Phase 2 project. Any data deposited into any SEQC2 data warehouses will be deemed as part of SEQC2 datasets;
2. No organization or individual participant may disseminate the SEQC2 datasets or SEQC2 data warehouses access information to others without written consent from the FDA/NCTR (Weida.Tong@fda.hhs.gov). Request for the SEQC2 data warehouses access shall be made to the FDA/NCTR (Weida.Tong@fda.hhs.gov and the designated working group leader);
3. Prior to acceptance for publication of the SEQC2 manuscripts, public presentation and/or publication of the SEQC2 data and any results generated from the data without the consent of the FDA/NCTR on behalf of the SEQC2 Consortium are prohibited. Analysis results based on the SEQC data must be shared with SEQC consortium;
4. You agree on the above data access policy automatically if downloading data from this ftp site.

FTP structure/organization: Similar to GIAB ftp



Index of <ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/>

 [Up to higher level directory](#)

Name

Size

Last Modified

 Confidentiality_Terms_for_Accessing_the_SEQC2_Datasets.txt	1 KB	9/25/2017 8:10:00 PM
 release		9/29/2017 1:58:00 PM
 technical		9/29/2017 5:45:00 PM
 data		

- Under ftp/technical: references used (NCI-GDC), method development data


Index of ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/technical/reference_genome/GRCh38/

 [Up to higher level directory](#)

Name

Size

Last Modified

 GRCh38.d1.vd1.fa	3078408 KB	4/26/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.amb	18 KB	4/27/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.ann	398 KB	4/27/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.bwt	3034743 KB	4/27/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.fai	127 KB	9/29/2017 7:55:00 PM
 GRCh38.d1.vd1.fa.pac	758686 KB	4/27/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.sa	1517372 KB	4/27/2016 12:00:00 AM
 GRCh38.d1.vd1.fa.tar.gz	854778 KB	9/29/2017 5:53:00 PM
 GRCh38.d1.vd1_BWA.tar.gz	3372964 KB	9/29/2017 6:05:00 PM
 readme	1 KB	9/29/2017 6:12:00 PM

- Under ftp/.data (dot for additional control, avoid bot downloaders)
<ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/.data/> (fastqs, vcfs etc)

Index of <ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/.data/>

 [Up to higher level directory](#)

Name

Size

Last Modified

 Chinese_Quartet		9/18/2017 6:53:00 PM
 HCC1395_cell_line		9/19/2017 3:07:00 PM
 HapMap_trio		10/2/2017 3:20:00 PM

Index of ftp://ftp-trace.ncbi.nlm.nih.gov/seqc/ftp/.data/HapMap_trio/fastqs/A01/

 [Up to higher level directory](#)

Name

Size

Last Modified


 lane1_read1.fastq.gz	47212267 KB	7/8/2017 1:33:00 PM
 lane1_read2.fastq.gz	50224271 KB	7/8/2017 3:55:00 PM













- Under ftp/release: official callsets with versions by samples

ABRF data with GIAB samples: Cornell University

- ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/use_cases/ABRF_NGS/

Index of /giab/ftp/use_cases/ABRF_NGS/project1_nist_pgp_dna/

 [\[parent directory\]](#)

Name	Size	Date Modified
 illumina_hiseq2500v4_baylor/		7/17/17, 11:26:00 AM
 illumina_hiseq2500v4_cornell/		7/13/17, 10:19:00 AM
 illumina_hiseq2500v4_urochester/		7/7/17, 4:09:00 PM
 illumina_hiseq4000_cornell/		7/7/17, 4:17:00 PM
 illumina_hiseq4000_mayo/		7/7/17, 8:02:00 PM
 illumina_hiseq4000_umichigan/		7/8/17, 6:40:00 AM
 illumina_hiseqx_broad/		7/6/17, 9:06:00 PM
 illumina_hiseqx_hudsonalpha/		7/10/17, 12:21:00 PM
 illumina_hiseqx_nygc/		7/10/17, 12:21:00 PM
 illumina_nextseq500_dfci/		7/10/17, 12:22:00 PM
 illumina_nextseq500_mbf/		7/10/17, 3:14:00 PM
 lifetech_proton_nci/		7/10/17, 12:23:00 PM
 lifetech_proton_uofu/		7/10/17, 3:20:00 PM

- ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/use_cases/epiqc/ for SEQC2 EpiQC WG